

# 高校学生学业成绩评定的现实、借鉴与实践<sup>\*</sup>

张劲英 孙 凯

**[摘要]** 学业成绩是高校学生学业成就评价的主要指标,也是高等教育评价的核心内容。科学的学生成绩评定对反映教学的真实情况、提高教学质量和促进教学改革至关重要。本文在把握相关研究与实践成果的基础上,针对我国高校学生成绩评定的现状,在观测点选择、数据分析方面提出了一些操作性较强的改进方案。

**[关键词]** 高等教育 学业成绩 教育评价

**[中图分类号]** G647 **[文献标识码]** A **[文章编号]** 1005—5843(2013)02—0011—04

**[作者简介]** 张劲英,厦门大学教育研究院与哥伦比亚大学教师学院联合培养博士生;孙凯,厦门大学教育研究院博士生(福建厦门 361005)

学生学习成效的评价对于高校检查和提高教学质量、优化学生管理以及推动教育教学改革都具有极其重要的意义。如何在教育部新一轮本科教学评估中加强对学生学习成效的关注,已成为高等教育理论与实践关注的热点。高校学生学习成绩是反映学生专业发展成果与潜力的一个重要指标,也是高校甄选学生、评定奖学金、授予学位等活动的重要依据,是高等学校教学和学生管理的重要内容。学业成绩评价是否科学、严谨,关乎高校教学与管理的权威性与公正性。然而,对比学业成绩所蕴含的丰富教育质量信息和实际利害关系,我国高等教育管理的理论研究与实践所做的探讨与分析并不多见。这固然与高校有关部门的管理模式和技术平台不一、信息共享和数据交换缺乏相应的规范与标准所造成的“信息孤岛”有关,而更重要的原因在于高校的教学、科研、管理活动对于如何从学业成绩的数据资源入手,考察学生的学术成长、提高教育教学质量管理,缺乏必要的理论与实践准备。

有鉴于此,本文在把握相关研究与实践发展成果的基础上,针对我国高校学业成绩评定的现实与问题,在观测点选择、数据分析等方面提出了一些操作性较强的方案,以期对我国当前高校的教育质

量管理相关研究与实践有所助益。

## 一、高校学生成绩的评价应用与適切性探讨:国外的经验

高等教育阶段的学生学业成绩评价主要依赖国家统一进行的标准化考试和高校校内考试系统。以美国高等教育的研究与实践为例,研究生入学考试等标准化学业考试成绩、就学期间常规学习的成绩表现——平均学分绩点等往往作为典型指标为研究者所采用。根据研究与实践的需要,研究者与管理部门可以计算学生的所有课程平均学分绩点、必修课程平均学分绩点或核心课程平均绩点等,以便更细致地了解学生的学业发展情况。

学习成绩相对于其他的学业表现反馈有便于量化、反馈及时、客观可靠、可比性强等优点<sup>①</sup>,其局限之处则在于考试成绩对短期记忆是很好的衡量工具,对长期的学习效果评价效果则未必理想。而就平均学分绩点这一指标而言,由于计入了学生多样化的课程成绩,平均学分绩点对学生的综合能力是一个很好的反映;局限之处也正在于平均学分绩点往往以学生的科目期末成绩为基础,反映的是学生课程学习的整体情况,但不能反映学生在特定学

<sup>\*</sup> 本文系教育部哲学社会科学重大课题攻关项目“我国本科人才培养质量研究”(项目编号:10JZD0038)的阶段性成果。

习任务中的表现。而且,成绩实际上是教师在特定环境下对学生表现的评价,具体的评价结果受教师的个人背景因素和环境因素(包括宏观层面的社会、院校、学科结构)的影响<sup>②</sup>。而具体到每一项课程考试成绩,影响因素则不仅有教师的授课质量、学生掌握知识的程度、考题的取样、命题的方法和难度等,还有考试的组织形式以及学生的临场发挥等诸多方面。

此外,当前大学生在校学习的成绩的确也面临一些公信力质疑,主要体现为研究者对“分数膨胀”、“分数压缩”和“成绩不对等”等问题的关注。分数膨胀即学习成绩的“水分”,有隐性和显性之分。隐性的分数膨胀指的是教师背景相似、背景和学业表现具有可比性的两组学生,后期的学生比早期的学生得到的成绩普遍较高的现象;显性的分数膨胀则是由于教师特点、院校政策等原因造成的。分数压缩是指分数在体现学生表现上的区分功能下降。当分数膨胀问题出现时,由于分数过于集中,不能很好地区分学生的表现。成绩不对等指的是不同学科、课程的打分标准存在实质性差异所带来的分数可比性问题<sup>③</sup>。

国外的实证研究中,分数不对等的的确是一个比较严重的问题。为此,研究者们需要借助计算标准分成绩等方法使成绩相对可比。高等教育的实际工作者们更关注的是分数膨胀和压缩所带来的学业成就评价权威性与公正性问题。美国的实证研究证实,院校层面的GPA的确有逐年上升的趋势,但分数膨胀的问题尚不严重。其中,显性的分数膨胀并不普遍,隐性的分数膨胀则相对得到了较多实证研究的确认。其原因在于,学生选课是推高学生个人和院校GPA的机制。虽然分数上升和分数膨胀会导致分数的区分度下降,但国外的文献回顾表明,实证研究尚未发现分数压缩的现象。当然,如果分数持续上升,分数压缩可能就不可避免了<sup>④</sup>。

## 二、我国高校学生成绩评价的实践与问题

我国高等教育阶段的学生学业成就评价也依赖标准化考试和高校校内考试系统。目前,我国高校学生参与程度较高的大规模标准化考试是大学英语考试。作为一项全国性的标准化英语水平综合测试,大学英语考试成绩的分数具有客观、权威等优点<sup>⑤</sup>,其局限在于考试的内容只是英语语言知识和语言技能,且基本形式为多项选择题,对学生语言综合运用能力和语言交际能力的考查相对不足<sup>⑥</sup>。

对学生大学英语考试成绩的考查,可以采取考试等级、考试成绩、考试成绩是否通过、是否首次参加考试即通过等观测指标。

受标准化考试的实施条件限制,校内纸笔考试、课程论文、课堂学习评价、专业实践(包括综合实践、实验、实习和实训)构成了学生学习评价的主要环节。校内考试是评价的核心,具有简单易行、省时经济、适应性强等特点,缺点则在于由于高校中许多课程没有统一的课程纲要,更没有学科教学质量评价指标体系,教师命题主观性很强,可能谈不上考题的信度、效度、难度和区分度。

每一种学习评价的指标都有其利弊优长。作为学生学习成效的关键量化指标,国外的研究与实际工作者对学习成绩指标的精确性与適切性进行了不断的推敲、验证。对比国外的经验,我国对于把考试成绩作为唯一评价标准和考试成绩评定偏颇之处的批评不绝于耳;与此同时,考试成绩作为高校甄选优秀学生、评定奖学金、授予学位等工作的重要依据,由于评价制度与技术都欠缺规范,其权威性和公正性都受到质疑。另外,由于不明晰指标的结构与功能,考试成绩本身也被寄予了过于理想化的预期。考试成绩是学习评价的“硬数据”,但它无法体现教育产出的其他重要表现形式,如学生的态度和价值观的转变等等。严格地讲,这也不是考试成绩作为学业成就评价指标的题中应有之义。构建有利于学生全面发展的学习评价体系是一个系统的工程,需要理论的积累、技术的普及以及大规模调查数据和实证研究的支持与验证,因而尚待时日。然而,基于已有的理论与实践积累,针对我们目前在学习成绩评定中的现实问题,尽量科学、严谨地量化分析学习成绩,是颇为自然的现实选择。因此,笔者将就我国高校学生平均学分绩点的计算与分析提出一些方法上的探讨,以期对评价的规范化有所助益。

## 三、高校学生平均学分绩点分析与计算方法探讨

### 1. 计分制度与分数计算

按照国外的高等教育传统,课程成绩评定较多采用五级制评分或多级制评分制度,并以相对应的数值计分,这种数值就称为绩点。学生的课程平均成绩也就相应按点计算,其他教学管理也都以此为基础。计算平均学分绩点一般是将学生某一教学阶段的每门课程的成绩乘以学分,加起来以后除以总

的学分得出平均分。我国高校主要采取百分制的课程计分方式。但自从 20 世纪 90 年代推行学分制以来, 高校基本上都采用绩点制作为其配套措施, 于是产生了课程考核成绩与绩点的转换问题。不少研究者围绕如何将百分制成绩、五级制成绩和二级制成绩转化成相应绩点的科学性和合理性进行探讨, 提出了不同的思路和方法。教学管理实践中多将百分制成绩硬性转换为绩点制, 采用的绩点换算方法不一, 有段——点型、点——点型、连续型。由于早期五点制绩点数之间跨度过大, 不易区分学生成绩的差别, 部分学校将学生成绩由 5 个分数段增加到 8 个或更多, 并增加相应的绩点等级<sup>⑦</sup>。也有研究者认为, 学分制客观上并不需要进行成绩与绩点的转换, 而且成绩与绩点转换虽然逐步得到细化, 但绩点制的精确度和区分度始终不如百分制, 因此这种转化总会产生误差, 所以以百分制直接计算平均学分绩点是最合理的。这种做法的局限在于在国际交流中可能会造成一些沟通的障碍<sup>⑧</sup>。我们在进行校内教学管理或者科研用途的成绩分析时, 可以不必顾虑这一局限。

除百分制之外, 国内高校往往还同时存在特殊计分制度——二级制、四级制的情况。处理采用特殊计分制度的课程成绩, 可以采取不同的方案。方案一是对特殊计分成绩按一定换算方法进行百分制换算。如某高校的学籍管理规定中的《考核与成绩记载》规定, 百分制与四级制的换算标准是: 85~100 分为优秀 (A); 70~84 分为良好 (B); 60~69 分为及格 (C); 未达 60 分为不及格 (F)。而该高校的教学管理部门在实际操作中, 将四级制成绩按上述标准换算成百分制, 按每一个四级制计分标准中所对应的最高百分制分数计入。两级制成绩则不计入加权平均成绩。

将特殊计分转化成百分制的做法并不符合统计科学。百分制计分对分数的测度比四级制和二级制更精确、高级。测度精确程度不同的数据无法直接合并计算。由于高精度的数据拥有低精度数据的全部特性, 我们可以将数据做精度由高到低的转换, 如将百分制计分转化为五分制, 但反之却不行。另外, 如果为了迁就特殊计分而将百分制成绩转化为四级、二级制计分, 虽然可行但将极大地损失数据的精确性。

另一方案是在不同分制的科目性质可以区分的情况下 (如所考查知识类型、考试的组织形式、命题的方法等), 将不同分制性质的成绩区别对待,

不但能满足成绩统计的精确性要求, 而且排除了分数差异的课程体制等因数。实际上, 高校课程中采取特殊计分方法的课程往往在教学环节、教学内容和模式、教学要求、成绩评定方式等方面也与一般的课堂教学科目有很大不同。因此某些学校在学籍管理中明确规定社会实践不计入平均学分, 或公共选修课不记入学籍。

平均学分绩点计算的一个重要问题在于原始的学习成绩具有不等值、无横向可比性和不能反应个体在团体中的相对位置等缺点, 于是有研究者提出了以标准分型绩点消除上述缺点的设计。成绩标准化处理过程如下:

首先计算课程平均分数

$$\bar{X}(s) = \frac{\sum_{i=1}^n X_i(s)}{n} \quad (1)$$

然后计算课程成绩标准差

$$\delta(s) = \sqrt{\frac{\sum_{i=1}^n [x_i(s) - \bar{X}(s)]^2}{n}} \quad (2)$$

最后得到标准分

$$Y_i = \frac{X_i(s) - \bar{X}(s)}{\delta_i} \quad (3)$$

其中,  $X_i(s)$  表示某学生第  $i$  门课程的考核成绩,  $X_i$  表示第  $i$  门课程的学分,  $\delta_i$  为第  $i$  门课程对应教学群体中全体学生的该课程考核成绩标准差,  $n$  表示选修  $i$  课程学生数。为了照顾计分习惯, 还可将标准分数还原成百分制分数或者绩点。经过标准化处理后的分数可以把不同教学群体间取得分数的差异在算法中得到体现, 使教学群体间的学生成绩比较成为可能。值得注意的是, 标准分绩点的具体计算方法假设成绩服从正态分布, 而现实中的课程成绩分布几乎不可能满足这样的理论要求。我们可以通过 SPSS、Excel 和 Stata 等统计软件绘制直方图或者密度曲线获得学生成绩的分布状态。如果学生成绩的分布与正态分布相差不大, 则可基本忽略由此引起的误差。如果实际成绩分布偏离正态分布程度较大, 在进行成绩标准化处理之前应先将其

正态化。数据正态化的方法，要视学生成绩的原始分布状态而定（笔者根据若干种偏态可能列出了处理非正态成绩的方法，如表 1 所示）。此外，考虑到不同课程的难易程度和重要性不同，还有研究者

提出引入难度及重要性修正系数对成绩加以区分。这种方法在原理和技术层面具有一定合理性，但在实际操作中很难界定不同课程的难度和重要性，因此实践层面的可行性并不高。

表 1 处理偏态成绩分布方法举例

数据原始状态	处理方法	公式
严重负偏态	立方	$new = old^3$
轻度负偏态	平方	$new = old^2$
严重正偏态	负倒数；平方根负倒数	$new = - (old^{-1})$ ; $new = - (old^{-0.5})$
轻度正偏态	自然对数；平方根	$new = \ln (old)$ ; $new = old^{0.5}$

注：new 指正态处理后的学生成绩；old 指原始学生成绩

## 2. 课程范围

选择哪些科目的成绩进入平均学分绩点的计算，实际操作中也有不同的做法。比如美国高中阶段平均学分绩点的计算，有些学校将全部课程都计入，有些学校则将非学术课程，如健康教育、体育、表演等排除在 GPA 的计算之外，只考虑“学术核心”课程的平均学分绩点。这种计算方式得到的学分绩点称作核心平均学分绩点。核心平均学分绩点主要包括五个领域的学术表现：数学、英语和语言艺术、社会科学、科学和外语。高选择性的大学在招生过程中，也会将 GPA 当中的“软信息”剥离，只考虑学生的核心学术能力。我国也有高校在学籍管理中明确规定军训、健康教育、体育课、任选课等教学环节的成绩不计入平均学分绩点。因此，笔者建议在计算 GPA 时只考察学生的智育课程成绩，而将体育、德育课程另作评价。此外，由于艺术、表演等艺术类专业高校学生在入学标准、课业学习方面均与其他学科、专业的学生有本质的区别，对他们的在校学业表现分析也应有所区别。

考虑到不同课程范围的成绩对学生学习能力和兴趣反应的综合程度不同，我们可以计算不同课程范围的学生课程 GPA。目前，高校在对学生进行与奖学金、研究生推荐等有关的成绩评价时会将所有学术课程的成绩计入，这样做的优点是能最大范围地衡量学生的学业表现与兴趣，但其中包含的选修课可能是学生不太重视的课程。必修课程是大多数学生都重视的课程，因此必修课程 GPA 能更集中体现学生的学习能力。另外，公共基本课程中的数学、英语反映的是学生的基本学习能力与兴趣，专业必修课程反映的是学生与专业相关的能力和兴趣。这两部分课程是学生学习能力与兴趣的最核心体现。

高校学生学业成绩的评定是一项重要的教学与管理环节，牵涉到学校教学质量评估与管理、学生评价与就业等工作的实施。因此，在研讨国内外高校学生学业成绩管理的基础上不断提高其科学性，既是高校教学、管理的实际问题，也是高等教育研究的重要课题。总的来说，清晰的评价目的、适当的指标和观测点选择、恰切的结果解释、合理的统计技术的采用是科学、严谨的学习成绩评定始终应遵循的原则。

注释：

①吴康宁. 教育社会学 [M]. 人民教育出版社, 1998: 238, 268.

②Shouping, Hu (Ed.) (2005). Beyond Grading Inflation: Problems in Higher Education. ASHE Higher Education Report, Vol 30, No. 6. San Francisco: Jossey Bass.

③Lavin, D. E. (1965). The Prediction of Academic Performance: A Theoretical Analysis and Review of Research, New York: Russell Sage Foundation.

④Mansfield, H. C (2001). Grade Inflation: It's time to face the facts. Chronicle of Higher Education, B24.

⑤杨惠中. 大学英语四、六级考试回顾与展望 [J]. 外语界, 1998, (3): 27.

⑥李炯英. 中国现行大学英语四、六级考试: 问题与思考——兼评国内外相关研究 [J]. 外语教学, 2002, (9).

⑦石曙光. 六种学分(绩点)制对比分析 [J]. 吉林教育科学·高教研究, 1997, (1).

⑧刘卫萍. 平均学分计算方法比较与选择 [J]. 当代教育论坛, 2008, (7): 32-33.

(责任编辑: 刘新才)